

# 表層類似度に基づくセンター試験『国語』現代文 傍線部問題ソルバー

佐藤 理史<sup>†</sup>・加納 隼人<sup>†</sup>・西村 翔平<sup>††</sup>・駒谷 和範<sup>†††</sup>

大学入試センター試験『国語』の現代文で出題される，いわゆる「傍線部問題」を解く方法を定式化し，実装した．本方法は，問題の本文の一部と5つの選択肢を照合し，表層的に最も類似した選択肢を選ぶことにより問題を解く．実装した方法は，「評論」の「傍線部問題」の半数以上に対して正解を出力した．

キーワード：大学入試センター試験，国語現代文，傍線部問題，表層類似度

## A Surface-Similarity-Based Solver of Comprehension Questions Referring to Underlined Passages in Contemporary Japanese of the National Center Test

SATOSHI SATO<sup>†</sup>, HAYATO KANOU<sup>†</sup>, SHOHEI NISHIMURA<sup>††</sup> and KAZUNORI KOMATANI<sup>†††</sup>

We formalized and implemented a method for solving comprehension questions referring to underlined passages in Contemporary Japanese of the National Center Test. A target question consists of a text body extracted from a critical essay or novel, a question sentence, and five choices; the question sentence refers to an underlined passage in the text body and asks a question related to the passage, such as the interpretation of the passage, the reason for the author's claim (in the essay), and the emotional state or feeling of a character (in the novel). Given a question, the method determines which text segment is the key to the question and selects the choice that is most similar to the segment. The method correctly solved more than half the “critical essay” questions in previous tests of the National Center Test.

**Key Words:** *the National Center Test, Contemporary Japanese, questions referring to underlined passages, surface similarity*

### 1 はじめに

日本において，大学入試問題は，学力（知力および知識力）を問う問題として定着している．この大学入試問題を計算機に解かせようという試みが，国立情報学研究所のグランドチャレンジ「ロボットは東大に入れるか」というプロジェクトとして2011年に開始された（新井，松崎

---

<sup>†</sup> 名古屋大学大学院工学研究科, Graduate School of Engineering, Nagoya University

<sup>††</sup> 名古屋大学工学部, School of Engineering, Nagoya University

<sup>†††</sup> 大阪大学産業科学研究所, The Institute of Scientific and Industrial Research, Osaka University

2012). このプロジェクトの中間目標は、2016 年までに大学入試センター試験で、東京大学の二次試験に進めるような高得点を取ることである。

我々は、このプロジェクトに参画し、2013 年度より、大学入試センター試験の『国語』現代文の問題を解くシステムの開発に取り組んでいる。次章で述べるように、『国語』の現代文の設問の過半は、傍線部問題とよばれる設問である。船口 (船口 1997) が暗に指摘しているように、『国語』の現代文の「攻略」の中心は、傍線部問題の「攻略」にある。

我々の知る限り、大学入試の『国語』の傍線部問題を計算機に解かせる試みは、これまでに存在しない<sup>1</sup>。そのため、この種の問題が、計算機にとってどの程度むずかしいものであるかさえ、不明である。このような状況においては、色々な方法を試すまえに、まずは、比較的単純な方法で、どのぐらいの正解率が得られるのかを明らかにしておくことが重要である。

本論文では、このような背景に基づいて実施した、表層的な手がかりに基づく解法の定式化・実装・評価について報告する。我々が実装したシステムの性能は、我々の当初の予想を大幅に上回り、「評論」の傍線部問題の約半分を正しく解くことができた。

以下、本稿は、次のように構成されている。まず、2 章で、大学入試センター試験の『国語』の構成と、それに含まれる傍線部問題について説明する。3 章では、我々が採用した定式化について述べ、4 章ではその実装について述べる。5 章では、実施した実験の結果を示し、その結果について検討する。最後に、6 章で結論を述べる。

## 2 センター試験『国語』と傍線部問題

大学入試センター試験の『国語』では、毎年、大問 4 題が出題される (教学社編集部 2013)。その大問構成を表 1 に示す。この表に示すように、現代文に関する出題は、第 1 問の「評論」と第 2 問の「小説」であり、『国語』の半分を占めている。

第 1 問の「評論」は、何らかの評論から抜き出された文章 (本文) と、それに対する 6 問の設問から構成される。6 問の内訳は、通常、以下のようになっている。

問 1 漢字の書き取り問題が 5 つ出題される。

表 1 センター試験『国語』の大問構成—出典 (教学社編集部 2013)

近代以降の文章 (現代文)	第 1 問	評論	50 点
	第 2 問	小説	50 点
古典	第 3 問	古文	50 点
	第 4 問	漢文	50 点

<sup>1</sup> CLEF2013 では、QA4MRE のサブタスクの一つとして、Entrance Exams が実施され、そこでは、センター試験の『英語』の問題が使用された。

問 2-問 5 本文中の傍線部について，その内容や理由が問われる。

問 6 本文全体にかかわる問題で，2006 年以降は本文の論の進め方や本文の構成上の特徴などが問われる。

一方，第 2 問の「小説」は，何らかの小説から抜き出された文章（本文）と，それに対する 6 問の設問から構成される。6 問の内訳は，通常，以下のようになっている。

問 1 語句の意味内容を問う問題が 3 つ出題される。

問 2-問 5 本文中の傍線部を参照し，登場人物の心情・人物像・行動の理由などが説明問題の形で問われる。

問 6 本文全体の趣旨や作者の意図，表現上の特徴などが問われる。

これらの設問のうち，「評論」「小説」の両者の問 2 から問 5 を傍線部問題と呼ぶ。傍線部問題の具体例を図 1 に示す。この図に示すように，傍線部問題は，原則として 5 つの選択肢の中から正解を一つ選ぶ 5 択問題である。傍線部問題の配点は，2009 年度本試験では，第 1 問 32 点，第 2 問 33 点の計 65 点であり，現代文の配点 100 点の約 2/3 を占める。

問 5 傍線部 D 「行為と行為をつなぐこの空間の密度を下げているのが，現在の住宅である」とあるが，それはどういうことか。その説明として最も適当なものを，次の 1～5 のうちから一つ選べ。

- 1 現在の住宅では，仕事部屋や子ども部屋など目的ごとに空間が切り分けられており，それぞれの用途とはかかわらない複数の異なる行為を同時に行ったり，他者との関係を作り出したりするような可能性が低下してしまっていること。
- 2 現在の住宅では，ゾーニングが普及することでそれぞれの空間の独立性が高められており，家族であってもそれぞれが自室で過ごす時間が増えることで，人と人が触れあい，関係を深めていくことが少なくなっていること。
- 3 現在の住宅では，空間の慣習的な使用規則に縛られない設計がなされており，居住者たちがそのときその場で思いついたことを実現できるように，各自がそれぞれの行為を同時に行えるようになっていること。
- 4 木造家屋などかつての居住空間では，居間や台所など空間ごとの特性が際立っていたが，現代の住宅では，居住者が部屋の用途を交換でき，空間それぞれの特性がなくなってきたこと。
- 5 木造家屋などかつての居住空間では，人体の運動と連動して空間が作り変えられるような特性があったが，空間ごとの役割を明確にした現在の住宅では，予想外の行為によって空間の用途を多様にすることが困難になっていること。

図 1 傍線部問題の例（2011 年度本試験第 1 問の問 5（2011M-E5））

表 2 使用する傍線部問題の一覧

	第 1 問「評論」	第 2 問「小説」
2001 本試験	問 2, 問 3, 問 4, 問 5	問 2, 問 3, 問 4, 問 5
追試験	問 2, 問 3, 問 4, 問 5	問 2, 問 3, 問 4, 問 5
2003 本試験	問 2, 問 3, 問 4, 問 5	問 2, 問 3, 問 4, 問 5
追試験	問 2, 問 3, 問 4, 問 5	問 2, 問 3, 問 4, 問 5
2005 本試験	問 2, 問 3, 問 4, 問 5	問 2, 問 3, 問 4, 問 5
追試験	問 2, 問 3, 問 4, 問 5	問 2, 問 3, 問 4, 問 5
2007 本試験	問 2, 問 3, 問 4, 問 5	問 2, 問 3, 問 4, 問 5
追試験	問 2, 問 3, 問 4, 問 5	—
2009 本試験	問 2, 問 3, 問 4, 問 5	問 2, 問 3, 問 4, 問 5
追試験	—	問 2, 問 3, 問 4
2011 本試験	問 2, 問 3, 問 4, 問 5	問 2, 問 3, 問 4
	計 40 問	計 38 問

## 使用する試験問題

本研究では、2001 年度から 2011 年度の奇数年の大学入試センター試験の本試験および追試験の『国語』（2005 年以前は『国語 I・II』）を使用する。ただし、諸般の事情により、本文等が欠けているものがあり、それらは使用しない。表 2 に、本研究で使用する傍線部問題の一覧を示す。

なお、以降では、設問を指し示す ID として、以下のような 4 つの情報を盛り込んだ形式を採用する。

- (1) 年度（4 桁）
- (2) 試験区分（M: 本試験, S: 追試験）
- (3) 出題区分（E: 評論, N: 小説）
- (4) 設問番号（2, 3, 4, 5）

たとえば、図 1 に示した、2011 年度本試験第 1 問「評論」の問 5 は、「2011M-E5」と表す。なお、この例に示したとおり、(2) と (3) の間に、ハイフン「-」を挟む。

## 3 傍線部問題の定式化

### 3.1 定式化

センター試験『国語』現代文傍線部問題では、正解となる選択肢の根拠が必ず本文中に存在すると指摘されている（船口 1997）。そして、その根拠となる部分は、傍線部の付近に存在することが多いことが、板野の分析によって明らかにされている（板野 2010）。我々は、これらの知見に基づき、傍線部問題を、図 2 に示すように定式化する。

この定式化は、次のことを意味する。

<b>入力</b>	(a) 本文 $T$
	(b) 設問文 $Q$ (本文中のどこを参照しているかの情報が含まれるものとする)
	(c) 選択肢集合 $C = \{c_1, c_2, c_3, c_4, c_5\}$
<b>出力</b>	$c = \arg \max_{c_i \in \text{pre\_select}(C, Q)} \text{score}(\text{extract}(T, Q), c_i) \text{polarity}(Q)$
<b>関数</b>	(a) $\text{extract}(T, Q)$ : 本文 $T$ から, その一部を抽出する
	(b) $\text{score}(\hat{T}, c_i)$ : テキスト $\hat{T}$ と $c_i$ から, 照合スコアを計算する
	(c) $\text{polarity}(Q)$ : 設問文の極性を判定する
	(d) $\text{pre\_select}(C, Q)$ : 選択肢集合 $C$ から, その部分集合を選ぶ

図 2 傍線部問題の定式化

- (1) それぞれの設問を入力とする。設問は, 本文  $T$ , 設問文  $Q$ , 選択肢集合  $C$  から構成されるものとする。(図 1 では, これらのうち, 本文  $T$  を除いた, 設問文  $Q$  と選択肢集合  $C$  を示している.)
- (2) 設問文  $Q$  と本文  $T$  から, 選択肢と照合する本文の一部  $\hat{T} = \text{extract}(T, Q)$  を定める。
- (3) 選択肢集合  $C$  の中から, 実際に本文の一部  $\hat{T}$  と照合する部分集合  $\hat{C} = \text{pre\_select}(C, Q)$  を定める。これは, 選択肢の事前選抜に相当する。
- (4) 設問文の極性  $\text{polarity}(Q)$  を判定する。この関数は, 設問文  $Q$  が「適当なもの」を要求している場合に  $+1$  を, 「適当でないもの」を要求している場合に  $-1$  を返すものとする。
- (5) 事前選抜済の選択肢集合  $\hat{C}$  に含まれる選択肢  $c_i$  と, 本文の一部  $\hat{T}$  との照合スコア  $\text{score}(\hat{T}, c_i)$  を計算する。これに設問文の極性  $\text{polarity}(Q)$  をかけたものを, その選択肢の最終スコアとする。
- (6) 事前選抜済の選択肢集合  $\hat{C}$  に含まれる選択肢  $c_i$  のなかで, 最終スコアが最大のものを, 解として出力する。

この定式化の特徴は, 図 2 に示した (a)–(d) の 4 つの関数に集約される。これらの背後にある考え方について, 以下で説明する。

### 3.2 本文の一部と照合する

正解の選択肢の根拠となる箇所は, 多くの場合, 本文中の傍線部の周辺にあると考えるのが妥当である。先に述べたように, この点についての詳細な分析が, 板野によってなされている(板野 2010)。実際, 我々人間が傍線部問題を解くとき, 本文中の傍線部の前後に注目するのは, 標準的な戦略である。

このような戦略は, 選択肢  $c_i$  を本文  $T$  全体と照合するのではなく, あらかじめ, 本文から,

根拠が書かれていそうな部分  $\hat{T}$  を抜き出し、 $\hat{T}$  と  $c_i$  の照合スコアを計算することで、具体化できる。

本文の一部を取り出す方法として、

- (1) 本文の先頭から、当該傍線部までを  $\hat{T}$  とする、
- (2) 一つ前の設問で参照された傍線部から、当該傍線部までを  $\hat{T}$  とする、
- (3) 傍線部の前後のある範囲を  $\hat{T}$  とする、

などの方法が考えられ、これらと、採用する単位（文または段落）を組み合わせることにより、多くのバリエーションが生まれることになる。もちろん、設問毎に、設問文  $Q$  や本文  $T$  に応じて異なる（適切な）方法を採用してもよい。

関数  $\text{extract}$  は、これらの方法を抽象化したもので、以下のような関数として定義する。

$$\hat{T} = \text{extract}(T, Q) \quad (1)$$

なお、厳密に言えば、設問文  $Q$  には一つ前の設問の傍線部の情報は含まれないが、その情報は本文を参照することによって得られるものと仮定する。事実、センター試験では、傍線部には A, B, C, D の記号が順に振られるため、当該傍線部が B であれば、一つ前の設問の傍線部は A であることがわかる。

### 3.3 照合スコアを採用する

本文の一部  $\hat{T}$  が、選択肢  $c_i$  とどの程度整合するか（その根拠となりうるか）を、照合スコア  $\text{score}(\hat{T}, c_i)$  として抽象化する。本来的には、整合するかないかの 2 値であるが、そのような判定を機械的に下すのは難しいので、0 から 1 の実数をとるものとする。

### 3.4 設問文の極性を考慮する

ほとんどの傍線部問題の設問文は、「最も適当なものを、次の 1～5 のうちから一つ選べ」という形式となっている。しかし、2001S-E4 のように、「適当でないものを、次の 1～5 のうちから一つ選べ」という形式も存在する。このような「適当でないもの」を選ぶ設問に対しては、照合スコアを逆転させる（照合スコアが最も小さなものを選択する）のが自然である。

上記のような設問形式に応じた選択法の変更を採用するために、設問文  $Q$  の極性を判定する関数  $\text{polarity}(Q)$  を導入する。この関数は、設問が「適当なもの」を要求している場合に +1 を、「適当でないもの」を要求している場合に -1 を返すものとする。

### 3.5 選択肢の事前選抜を導入する

本定式化では, 「設問文と 5 つの選択肢をよく読めば, 本文を参照せずとも, 正解にはならない選択肢のいくつかをあらかじめ排除できる」場合が存在すると考え<sup>2</sup>, 選択肢の事前選抜を明示的に導入する. 事前選抜 `pre_select` は, 選択肢集合  $C$  と設問文  $Q$  から,  $C$  の部分集合を返す関数として定式化する.

$$\hat{C} = \text{pre\_select}(C, Q), \quad \hat{C} \subseteq C \quad (2)$$

## 4 実装

前節の定式化に基づいて傍線部問題ソルバーを実装するためには, `score`, `polarity`, `extract`, `pre_select` の 4 つの関数を実装する必要がある. 表 3 に, 今回実装した方法の概要を示す (詳細は, 以下で説明する). なお, 前節の定式化では, 正解と考えられる選択肢を一つ出力する形になっているが, 実際のシステムは, 選択肢を照合スコア順にソートした結果 (すなわち, それぞれの選択肢の順位) を出力する仕様となっている. なお, 照合スコアが一致した場合は, 選択肢番号の若いものを上位とする<sup>3</sup>.

表 3 実装の概要

$\text{score}(\hat{T}, c_i)$	$A$	$\text{overlap\_ratio}_D(A; \hat{T}, c_i)$ (文字オーバーラップ率)
	$A^2$	$\text{overlap\_ratio}_D(A^2; \hat{T}, c_i)$ (文字 bigram のオーバーラップ率)
	$W$	$\text{overlap\_ratio}_D(W; \hat{T}, c_i)$ (形態素表層形のオーバーラップ率)
	$L$	$\text{overlap\_ratio}_D(L; \hat{T}, c_i)$ (形態素原形のオーバーラップ率)
$\text{polarity}(Q)$		表層の手がかりに基づき, 極性判定を行なう
$\text{pre\_select}(C, Q)$	no	選択肢の事前選抜を行なわない
	yes	選択肢のうちの 1 つを除外する
$\text{extract}(T, Q)$	P-a-e	本文すべて
	P-a-0	本文先頭から, 当該傍線部を含む段落まで
	P-b-0	ひとつ前の傍線部を含む段落から, 当該傍線部を含む段落まで
	P-b-e	ひとつ前の傍線部を含む段落から, 本文末尾まで
	P-m-n	当該傍線部を含む段落と, その前 $m$ 段落, 後 $n$ 段落
	S-m-n	当該傍線部を含む文と, その前 $m$ 文, 後 $n$ 文
	$\bar{X}$ (bar)	(抽出法 $X$ で, 当該傍線部を含む文を除外)

<sup>2</sup> 実際にセンター試験を受験したことがある複数人の意見に基づく.

<sup>3</sup> 結果に再現性をもたせるために, 照合スコアが一致したもののからランダムに選ぶ方法は採用しない.

## 4.1 オーバーラップ率

照合スコア score, および, 選択肢の事前選抜 pre\_select の実装には, オーバーラップ率を用いる. オーバーラップ率の定義には, 服部と佐藤の定式化 (服部, 佐藤 2013; Hattori and Sato 2013) を採用する.

この定式化では, まず, ある集合  $E$  を仮定する. この集合の要素が, オーバーラップ率を計算する際の基本単位となる. 集合  $E$  としては, たとえば, 文字集合  $A$ , 形態素集合  $W$ , あるいは, 文字  $n$ -gram の集合  $A^n$ などを想定する.

オーバーラップ率の算出の出発点となる式は, 2つの文字列  $t_1$  と  $t_2$  に共通に出現する集合  $E$  の要素の数を求める次式である.

$$\text{overlap}(E; t_1, t_2) = \sum_{e \in E} \min(\text{fr}(e, t_1), \text{fr}(e, t_2)) \quad (3)$$

ここで,  $\text{fr}(e, t)$  は, 文字列  $t$  における  $e (e \in E)$  の出現回数を表す. この値を,  $t_2$  の長さ, あるいは,  $t_1$  と  $t_2$  の長さの和で正規化することにより, オーバーラップ率を定義する.

$$\text{overlap\_ratio}_D(E; t_1, t_2) = \frac{\text{overlap}(E; t_1, t_2)}{\sum_{e \in E} \text{fr}(e, t_2)} \quad (4)$$

$$\text{overlap\_ratio}_B(E; t_1, t_2) = \frac{2 \text{overlap}(E; t_1, t_2)}{\sum_{e \in E} \text{fr}(e, t_1) + \sum_{e \in E} \text{fr}(e, t_2)} \quad (5)$$

前者の  $\text{overlap\_ratio}_D$  は,  $t_2$  の長さのみで正規化したもので, 方向性を持った (directional) オーバーラップ率となる. 後者の  $\text{overlap\_ratio}_B$  は,  $t_1$  と  $t_2$  の長さの和で正規化したもので, 方向性を持たない, 双方向性 (bidirectional) のオーバーラップ率となる.

## 4.2 照合スコア

本文の一部  $\hat{T}$  と選択肢  $c_i$  の照合スコアには, 方向性を持ったオーバーラップ率  $\text{overlap\_ratio}_D$  を用いる.

$$\text{score}(\hat{T}, c_i) = \text{overlap\_ratio}_D(E, \hat{T}, c_i) \quad (6)$$

ここで, オーバーラップを測る際の単位 (要素) 集合  $E$  として, 以下の4種類を実装した.

- (1)  $A$ : 文字集合
- (2)  $A^2$ : 文字 bigram の集合
- (3)  $W$ : 形態素表層形の集合
- (4)  $L$ : 形態素原形の集合



いずれの場合も, 句読点は要素に含めなかった.

形態素解析器には mecab-0.994 を, 形態素解析辞書には, ipadic-2.7.0 または unidic-2.1.0 を用いた. すなわち,  $W$  と  $L$  は, それぞれ 2 種類存在することになる.

### 4.3 設問文の極性判定

設問文の極性判定は, 文字列マッチングで実装した. 具体的には, 正規表現「/(適切|適当)でないものを/」に一致した場合は negative (-1), それ以外は positive (+1) と判定する. 対象とした問題は限られているので, 極性判定結果は, 人間の判断とすべて一致する.

### 4.4 本文の一部の抽出

段落 (P) 単位および文 (S) 単位の抽出を実装した. 抽出する領域は, 連続領域を採用した. すなわち, 抽出単位, 抽出開始点, 抽出終了点の 3 つの情報によって, 抽出領域は定まる.

抽出開始・終了点は, 当該傍線部を含む単位 (段落または文) を基準点 0 とし, その前後何単位であるかを, 整数で表す. たとえば, S- $m$ - $n$  は, 当該傍線部を含む文と, その前  $m$  文, 後  $n$  文を表す (全部で  $m+1+n$  文となる). この他に, 本文先頭 (a), 前問の傍線部の位置 (b), 本文末尾 (e) という 3 種類の特別な位置を指定できるようにした. さらに, 当該傍線部を含む文を除外するというオプション ( $\overline{X}$ ) も実装した<sup>4</sup>.

### 4.5 選択肢の事前選抜

選択肢の事前選抜には, 次の方法を採用した.

- (1) それぞれの選択肢  $c_i$  において, 以下に示す事前選択スコア  $ps(c_i, C)$  を計算する.

$$ps(c_i, C) = \frac{1}{|C|-1} \sum_{c_j \in C, c_j \neq c_i} \text{overlap\_ratio}_B(A; c_i, c_j) \quad (7)$$

このスコアは, 他の選択肢  $c_j$  との双方向文字オーバーラップ率  $\text{overlap\_ratio}_B(A; c_i, c_j)$  の平均値である.

- (2) 得られた事前選択スコアが低い選択肢を, 選択肢集合から除外する. (最終順位付けでは, かならず 5 位とする)

なお, この実装は, いわば「もっとも仲間はずれの選択肢を一つ除外する」という考え方に基づいている.

<sup>4</sup> 設問の多くは, 傍線部のある種の言い換えを求めているので, 傍線部自身は, 選択肢を選ぶ根拠とはならないことが多いと考えられる. 今回の実装では, テキストを扱う最小単位は文なので, 「当該傍線部を含む文」を除外するという実装となった.

## 5 実験と検討

### 5.1 実験結果

実装した傍線部問題ソルバーを用いて、評論傍線部問題 40 問を解いた結果を表 4 および表 5 に示す。この表の各行の先頭の欄 (ID) は、本文抽出法 (extract) に対応しており、次の 2 つの数字は、その抽出法 (ID) で抽出された文数 (40 問の平均値)、および、該当傍線部を含む文を除外した場合 ( $\overline{\text{ID}}$ ) の文数を示す<sup>5</sup>。斜線で区切られた 4 つの数字は、ある要素集合を単位としてオーバーラップ率を計算した場合に対応し、それぞれの数字は、順に、以下の場合の正解数を示す。

- (1) 抽出法 ID + 事前選択なし (no)
- (2) 抽出法  $\overline{\text{ID}}$  + 事前選択なし (no)
- (3) 抽出法 ID + 事前選択あり (yes)
- (4) 抽出法  $\overline{\text{ID}}$  + 事前選択あり (yes)

表 4 の 2 行目 (P-a-0) の A 欄の最初の数字 20 が、我々に衝撃を与えた数字である。これは、本文の先頭から当該傍線部を含む段落までを  $\hat{T}$  として抽出し (P-a-0),  $\hat{T}$  と各選択肢  $c_i$  との照合スコアを文字オーバーラップ率 ( $A$ ) で計算して、スコアが最大値を取る選択肢を選んだ場合、

「評論」の傍線部問題の半分 (20/40) が正しく解ける

ことを意味する。

センター試験の設問は 5 択問題であるので、解答する選択肢をランダムに選んだとしても 1/5 の確率で正解する。40 問においてランダムに解答を選んだ場合、正解する問題数は、 $8 \pm 4.96$  ( $p = 0.05$ ) である<sup>6</sup>。この値と比べ、正解数 20 問は有意に多い。

我々は、このような性能が得られることを、まったく予期していなかった。この結果を受けて、我々は、色々な設定 ( $84 \times 6 \times 4 - 12 = 2,004$  通り)<sup>7</sup>での性能を網羅的に調べた。こうして得られたのが表 4 と表 5 である。これらの表では、正解数 20 以上をボールド体で表示した。さらに、正解数が 22 以上となった 16 の設定とその設定における正解の順位分布 (第  $n$  位として出力された正解がいくつあるか) を、表 6 に示した。

<sup>5</sup> S- $m$ - $n$  で文数が  $m+1+n$  を越えるのは、2003S-E5 の設問文が複数の傍線部 (正確には、波線部) を含むためである。この場合、最初に現れる波線部の前方  $m$  文から、最後に現れる波線部の後方  $n$  文までを抽出する。

<sup>6</sup>  $B\left(40, \frac{1}{5}\right) \approx N\left(8, 2.53^2\right)$ 。故に  $1.96 \times 2.53 \approx 4.96$

<sup>7</sup> 傍線部を含む段落が 1 文のみから構成されている場合があるので、抽出法  $\overline{\text{P-0-0}}$  は設定しない。これが -12 に相当する。84 は、表 4-5 の行数の合計を、 $6 \times 4$  は 1 行に記述される設定数を示す。

表 4 「評論」に対する実験結果 (その 1)

ID	ave. s.	A		A <sup>2</sup>		W-ipadic		W-unidic		L-ipadic		L-unidic	
	ID/ $\overline{\text{ID}}$	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
P-a-e	68.9/67.8	14/14/18/18		14/12/16/14		15/17/17/17		14/15/16/16		15/15/16/16		13/13/15/15	
P-a-o	40.0/39.0	<b>20/21/21/22</b>		15/13/17/15		14/14/15/15		11/13/12/13		13/15/15/16		11/12/12/13	
P-b-o	21.5/20.5	19/19/ <b>22/22</b>		14/13/16/16		13/13/15/16		11/12/12/14		14/16/17/18		12/14/13/15	
P-b-e	50.4/49.3	14/14/18/18		13/12/15/14		13/14/16/16		13/14/16/16		14/14/16/16		10/10/13/13	
P-0-0	6.0/ -	14/ -/17/ -		12/ -/17/ -		12/ -/15/ -		11/ -/14/ -		8/ -/13/ -		13/ -/16/ -	
P-1-0	10.8/ 9.7	16/17/19/19		14/14/16/17		12/15/17/18		15/17/17/18		14/15/18/17		13/15/14/16	
P-1-1	15.2/14.2	15/16/16/16		14/16/15/18		15/15/17/17		16/18/17/18		18/17/19/18		13/13/14/15	
P-2-0	15.2/14.2	18/18/ <b>21/21</b>		15/14/17/18		12/13/16/16		13/13/15/15		12/14/15/16		13/13/15/15	
P-2-1	19.6/18.6	17/17/18/18		12/14/13/15		14/14/16/16		14/13/16/15		17/16/19/17		13/13/15/15	
P-2-2	24.2/23.2	16/15/19/18		16/16/16/16		17/17/19/19		19/19/ <b>21/21</b>		19/19/ <b>20/19</b>		16/15/18/17	
P-3-0	19.1/18.1	18/16/ <b>21/19</b>		15/16/17/18		11/12/14/14		13/14/15/16		12/14/14/15		10/12/13/15	
P-3-1	23.6/22.5	18/16/19/17		11/12/12/13		11/10/12/11		12/12/13/13		14/14/16/15		10/10/12/12	
P-3-2	28.1/27.1	17/16/ <b>20/18</b>		13/14/14/15		14/13/15/14		16/16/17/17		16/15/17/16		13/13/15/15	
P-3-3	31.3/30.3	15/14/18/16		12/12/14/14		11/12/13/13		14/14/15/15		14/14/14/14		13/12/15/14	
P-4-0	22.8/21.8	18/16/ <b>21/18</b>		15/14/17/16		9/10/11/11		10/12/12/14		11/13/13/15		13/14/15/16	
P-4-1	27.2/26.2	17/16/19/17		9/ 9/10/10		12/12/13/12		11/12/11/12		14/13/15/14		12/12/14/14	
P-4-2	31.9/30.8	16/14/19/17		12/12/12/12		15/14/15/14		15/16/16/17		17/16/17/16		13/13/15/15	
P-4-3	35.0/34.0	14/13/19/17		11/10/12/11		12/13/13/13		13/14/14/15		15/14/15/14		13/13/15/15	
P-4-4	38.2/37.1	14/14/17/16		10/10/11/10		14/14/14/14		15/16/15/16		17/16/17/16		14/14/16/16	
S-1-0	2.3/ 1.2	10/11/14/13		11/15/16/17		10/ 9/16/14		10/11/13/12		6/11/11/14		9/11/12/12	
S-1-1	3.3/ 2.2	12/14/14/15		10/10/14/13		10/11/14/14		11/13/14/16		11/ 8/15/12		9/ 8/11/12	
S-2-0	3.3/ 2.2	16/11/ <b>20/15</b>		14/13/17/18		13/11/16/12		12/11/15/12		12/13/16/16		11/11/12/15	
S-2-1	4.3/ 3.2	14/14/15/17		12/14/15/16		16/13/17/14		11/13/13/15		10/11/13/13		12/12/14/15	
S-2-2	5.2/ 4.2	14/14/15/16		12/15/14/17		14/13/16/14		14/15/15/16		13/12/16/14		14/12/16/14	
S-3-0	4.3/ 3.2	12/11/16/15		16/18/18/ <b>20</b>		17/16/19/19		14/14/15/15		12/15/16/18		13/13/14/16	
S-3-1	5.3/ 4.2	14/15/16/17		15/16/16/19		16/14/18/16		12/12/15/15		10/11/13/14		15/12/17/15	
S-3-2	6.2/ 5.2	16/14/18/17		15/17/16/19		16/16/18/18		15/16/16/18		13/12/16/15		14/14/15/15	
S-3-3	7.1/ 6.0	12/14/14/15		16/15/17/16		16/16/17/17		14/15/14/16		16/14/18/15		14/15/15/16	
S-4-0	5.3/ 4.2	13/13/18/19		15/14/18/ <b>20</b>		17/14/ <b>20/18</b>		14/12/17/15		10/12/13/17		16/15/17/18	
S-4-1	6.3/ 5.2	16/17/18/ <b>20</b>		12/16/15/18		14/13/18/17		15/14/18/16		12/13/17/16		15/17/18/17	
S-4-2	7.2/ 6.2	17/18/19/ <b>20</b>		11/12/13/14		14/14/18/17		15/16/17/17		13/12/18/17		14/15/16/17	
S-4-3	8.1/ 7.0	15/17/17/19		13/14/15/16		14/15/17/17		14/15/15/17		15/14/19/17		14/16/16/18	
S-4-4	8.9/ 7.9	16/15/18/16		13/14/15/16		16/15/19/18		13/14/13/14		17/16/ <b>20/18</b>		14/14/16/16	
S-5-0	6.2/ 5.2	13/15/18/ <b>21</b>		13/15/18/ <b>20</b>		15/13/19/17		15/16/18/16		12/13/16/16		15/15/17/17	
S-5-1	7.2/ 6.2	15/17/17/ <b>20</b>		13/17/16/ <b>20</b>		13/15/16/18		15/18/18/18		12/13/17/16		12/16/15/16	
S-5-2	8.2/ 7.1	16/18/18/ <b>20</b>		12/14/14/17		15/14/19/18		16/16/19/19		13/12/17/16		14/18/16/19	
S-5-3	9.1/ 8.0	15/16/16/17		12/14/14/16		15/15/19/17		14/16/16/18		14/14/18/17		15/17/17/19	
S-5-4	9.9/ 8.8	15/16/17/17		11/13/14/15		15/16/18/19		13/15/14/16		14/13/17/16		14/14/16/16	
S-5-5	10.7/ 9.7	16/17/18/18		11/14/13/15		14/14/16/16		13/14/14/15		15/14/18/17		13/13/15/15	
S-6-0	7.2/ 6.1	14/18/19/ <b>22</b>		17/18/ <b>20/22</b>		18/17/ <b>21/20</b>		17/17/19/18		15/16/18/19		18/15/19/18	
S-6-1	8.2/ 7.1	16/18/18/ <b>20</b>		15/19/18/ <b>22</b>		16/17/19/19		15/18/17/19		15/16/19/19		14/17/17/18	
S-6-2	9.1/ 8.1	17/ <b>20/18/21</b>		12/14/15/16		15/16/18/19		16/16/17/17		16/16/19/18		16/19/17/ <b>20</b>	
S-6-3	10.0/ 8.9	17/ <b>20/19/20</b>		12/14/15/16		14/16/17/18		13/15/15/17		15/15/18/16		14/15/15/16	
S-6-4	10.8/ 9.8	16/18/17/18		12/13/14/14		15/16/17/18		12/15/12/15		16/16/17/17		13/14/14/15	
S-6-5	11.7/10.6	16/17/17/18		12/13/13/14		15/14/16/15		11/13/11/13		16/15/17/16		13/14/14/15	
S-6-6	12.5/11.4	16/18/17/18		13/13/14/15		14/14/16/16		12/13/12/13		16/16/19/19		13/13/16/16	

表 5 「評論」に対する実験結果 (その 2)

ID	ave. s. ID/ $\overline{ID}$	A		$A^2$		W-ipadic		W-unidic		L-ipadic		L-unidic	
		no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
S-7-0	8.1/ 7.0	15/18/19/21		16/18/20/22		15/16/19/19		16/17/20/19		13/15/17/19		16/13/19/17	
S-7-1	9.1/ 8.1	19/19/20/20		14/19/17/22		15/18/18/19		15/18/17/19		14/17/17/19		14/16/17/18	
S-7-2	10.0/ 9.0	19/21/19/21		12/14/16/17		15/16/18/18		15/15/17/17		16/15/19/18		15/18/18/21	
S-7-3	10.9/ 9.8	18/21/19/20		12/14/15/16		15/16/18/17		13/15/15/17		16/15/19/17		15/16/16/17	
S-7-4	11.8/10.7	19/19/19/18		14/15/16/17		16/16/18/17		11/13/12/14		16/15/18/16		13/13/14/14	
S-7-5	12.6/11.5	18/17/18/17		15/14/16/15		16/15/17/16		11/12/11/12		16/15/18/16		13/13/14/14	
S-7-6	13.4/12.3	19/17/18/17		14/14/15/16		15/16/16/17		12/13/12/13		16/16/20/19		14/14/17/17	
S-7-7	14.2/13.2	19/18/19/18		16/16/16/16		16/16/18/18		13/14/13/14		17/17/21/20		13/13/17/17	
S-8-0	9.0/ 8.0	16/15/20/19		15/17/21/22		16/15/20/18		18/18/22/19		15/17/18/21		19/15/20/18	
S-8-1	10.0/ 9.0	19/17/20/18		14/19/19/22		16/18/18/18		18/19/19/18		15/17/19/20		16/18/19/20	
S-8-2	10.9/ 9.9	20/19/20/20		15/17/18/18		16/18/19/19		16/17/17/18		16/17/20/20		15/18/18/21	
S-8-3	11.8/10.8	20/20/20/19		13/16/15/18		15/16/18/17		13/16/15/18		15/16/18/17		15/17/16/18	
S-8-4	12.7/11.6	19/18/19/17		15/17/17/19		16/17/19/20		13/16/13/16		16/17/18/18		14/14/15/15	
S-8-5	13.5/12.4	17/16/17/16		16/16/17/17		15/16/18/19		12/14/13/15		16/17/18/18		14/14/15/15	
S-8-6	14.3/13.3	20/18/19/18		14/14/15/15		14/16/17/18		12/14/13/15		16/17/20/20		14/14/16/16	
S-8-7	15.2/14.1	20/19/20/19		15/15/16/16		15/16/19/19		13/15/14/16		17/18/21/21		13/13/16/16	
S-8-8	16.0/14.9	19/18/20/18		15/16/16/17		18/18/21/21		15/17/16/18		19/19/22/22		14/14/17/17	
S-9-0	9.9/ 8.9	17/16/21/20		16/18/20/23		15/16/20/20		15/17/19/18		17/19/19/23		18/15/18/17	
S-9-1	10.9/ 9.9	18/17/20/18		14/17/19/21		14/17/17/19		15/17/17/17		14/18/18/21		14/16/17/19	
S-9-2	11.8/10.8	20/20/20/19		16/17/19/19		14/16/18/18		15/16/17/17		14/16/18/19		14/15/17/18	
S-9-3	12.7/11.7	19/18/19/17		13/15/15/17		13/15/17/16		14/15/17/17		12/13/15/14		15/15/16/16	
S-9-4	13.6/12.5	18/18/19/18		14/16/16/17		14/15/17/18		14/15/15/15		13/14/15/15		13/13/14/14	
S-9-5	14.4/13.3	17/17/17/16		14/14/15/15		13/14/16/17		13/14/15/15		13/14/15/15		13/13/14/14	
S-9-6	15.2/14.2	20/19/20/19		13/13/14/14		13/13/16/16		12/14/14/15		13/14/17/17		14/14/16/16	
S-9-7	16.1/15.0	19/19/20/19		14/14/15/15		13/14/17/17		13/14/15/15		16/17/20/20		14/14/16/16	
S-9-8	16.9/15.8	18/18/19/18		15/15/16/16		16/17/19/19		15/16/17/17		18/18/21/21		15/15/17/17	
S-9-9	17.7/16.6	18/18/19/18		15/16/15/16		17/19/20/20		17/18/20/20		19/20/21/22		16/16/19/19	
S-10-0	10.8/ 9.8	17/16/21/20		15/16/19/21		14/14/18/17		15/18/19/19		15/16/17/19		16/15/18/18	
S-10-1	11.8/10.8	18/17/19/18		14/17/19/21		13/15/15/17		14/17/16/17		15/17/19/20		15/16/18/19	
S-10-2	12.8/11.7	20/20/20/19		14/16/17/18		13/14/16/16		14/16/16/17		16/16/20/19		16/16/19/19	
S-10-3	13.6/12.6	20/19/20/18		13/15/15/17		12/13/15/14		13/15/16/17		15/15/18/16		16/16/17/17	
S-10-4	14.5/13.4	19/19/19/18		13/15/15/16		14/15/17/18		14/16/15/16		14/14/16/15		15/15/15/15	
S-10-5	15.3/14.2	18/18/18/17		13/14/14/15		13/14/16/17		13/15/15/16		14/14/16/15		15/15/15/15	
S-10-6	16.1/15.1	21/20/21/20		13/14/15/16		13/13/16/16		13/16/15/17		14/14/17/16		16/16/17/17	
S-10-7	16.9/15.9	19/19/21/20		15/16/15/16		13/14/17/17		14/16/16/17		17/17/20/19		16/16/18/18	
S-10-8	17.8/16.7	18/18/20/19		16/17/16/17		16/17/19/19		14/16/16/17		18/18/21/20		16/16/18/18	
S-10-9	18.6/17.5	18/18/20/19		17/18/16/17		16/18/19/19		16/18/19/20		19/20/21/21		18/18/20/20	
S-10-10	19.4/18.3	17/17/19/18		15/17/16/16		15/15/18/18		16/18/19/20		18/19/20/20		17/17/19/19	

## 5.2 実験結果を検討する

表 4 と表 5 を観察すると、以下のことに気づく。

- (1) 照合するテキスト  $\hat{T}$  が極端に短い場合を除き、ほとんどの場合 (2,004 通り中 1,828 通りの設定) で、正解数はランダムな方法より有意に多い。すなわち、「評論」の傍線部問題に対しては、本論文で示した解法は、有効に機能する。
- (2) 照合スコアのオーバーラップ率の計算には、文字 (A) を用いると相対的に成績がよい場合が多い。文字オーバーラップ率が有効に機能するという、この結果は、日本語の含意

表 6 正解数が 22 以上の設定と正解の順位分布 (「評論」)

ID	bar	score	ps	ave.	1	2	3	4	5
S-9-0	bar	$A^2$	yes	8.9	23	8	2	6	1
S-9-0	bar	$L$ -ipadic	yes	8.9	23	6	2	8	1
P-a-0	bar	$A$	yes	39.0	22	5	7	5	1
P-b-0		$A$	yes	21.5	22	6	5	6	1
P-b-0	bar	$A$	yes	20.5	22	6	6	5	1
S-6-0	bar	$A$	yes	6.1	22	5	6	6	1
S-6-0	bar	$A^2$	yes	6.1	22	9	3	5	1
S-6-1	bar	$A^2$	yes	7.1	22	8	4	5	1
S-7-0	bar	$A^2$	yes	7.0	22	8	2	7	1
S-7-1	bar	$A^2$	yes	8.1	22	6	5	6	1
S-8-0	bar	$A^2$	yes	8.0	22	8	2	7	1
S-8-0		$W$ -unidic	yes	9.0	22	5	5	7	1
S-8-1	bar	$A^2$	yes	9.0	22	6	5	6	1
S-8-8		$L$ -ipadic	yes	16.0	22	3	5	9	1
S-8-8	bar	$L$ -ipadic	yes	14.9	22	5	3	9	1
S-9-9	bar	$L$ -ipadic	yes	16.6	22	7	3	7	1

認識 (RITE2) における服部らの結果 (服部, 佐藤 2013; Hattori and Sato 2013) に合致する. 文字オーバーラップ率が有効に機能するのは, おそらく, 日本語の文字の種類が多いこと, および, 漢字 1 文字が内容的情報を表しうること, の 2 つの理由によるものと考えられる.

- (3) 照合スコアのオーバーラップ率の計算に, 文字 bigram ( $A^2$ ), 形態素出現形 ( $W$ ), 形態素原形 ( $L$ ) を用いた場合は, 比較的短い  $\hat{T}$  のいくつかに対して, 成績がよい. これは, 比較的短いテキストの照合では, 語や文字 bigram などの, より長い要素の一致が大きな意味を持つためと考えられる. 今回の実験で最も成績がよかった正解数 23 は, 抽出法  $\overline{S-9-0}$ , 照合法  $A^2$  または  $L$ -unidic, 事前選抜あり (yes) の場合に得られた.
- (4) 照合テキスト  $\hat{T}$  から, 当該傍線部を含む文を除外した方が, 除外しなかった場合よりも, 成績は若干よい傾向を示す. 脚注 4 でも述べたように, 傍線部の言い換えを求めるような設問では, 該当傍線部自身は, 選択肢の根拠とはならないことが多い. このような設問に対しては, 該当傍線部を含む文を除外することによる効果があると考えられる.
- (5) 選択肢の事前選抜は, 正解数を増やす効果が見られる. なお, 今回使用した 40 問において, 正解が事前選抜によって除外される設問は, 1 問 (2005M-3) だけ存在した.
- (6) 用いる形態素解析辞書によって, 得られる結果は若干異なる. これは, 形態素として認

定する単位, および, 原形の認定法の違い<sup>8</sup>による. 今回の実験では, ipadic を使用した方が, 相対的によい結果が得られた場合が多かった.

### 5.3 性能の上限を見積もる

本論文で提案した方法で, どの程度の性能が達成可能であるかを見積もってみよう. 性能の上限は, それぞれの設問において,

- (1) 最も適切な本文の一部  $\hat{T}$  が選択でき, かつ,
- (2) 最も適切な照合スコアを選択できる

と仮定した場合の正解率で与えられる. ここでは,

- 形態素解析辞書には ipadic のみを用いる
- 選択肢の事前選択は行なわない<sup>9</sup>

こととした 668 ( $= 84 \times 4 - 4$ ) 通りの設定<sup>10</sup>を採用し, 各設問毎に 668 通りの設定の成績 (正解の順位) を集計した. その結果を表 7 に示す.

表 7 に示すように, 668 通りの設定のいずれにおいても正解を出力できなかった設問は, 2 問 (2001S-E5 と 2009M-E4) のみであった. すなわち,  $38/40 (= 95\%)$  の設問に対して, 本論文で示した解法は正解を出力できる可能性がある.

表 7 において, 1 位となった設定数が多い設問は, 言わば「ストライクゾーンが広い」設問である. つまり, パラメータ選択に「鈍感」であり, 機械にとってやさしい設問である. たとえば, 2003S-E3 (1 行目) は, 668 通り中 664 通りの設定で正解が得られている. 逆に, 1 位となった設定数が少ない設問は, 正解を出力するのが難しい設問である. たとえば, 2003M-E5 (下から 3 行目) は, 668 通り中 14 通りの設定でしか正解が得られない.

これらのことを考慮して, 次に, もうすこし現実的な到達目標を考えよう. 正解率  $1/5$  でランダムに 668 回の試行を行なった場合の正解数は  $133.6 \pm 20.26$  ( $p = 0.05$ ) である. この値の上限をひとつの目安として<sup>11</sup>, これよりも多い正解数が得られた設問は, 設問に応じた適切なパラメータ選択により, 正解を導ける可能性が高いとみなそう. このような設問は, 40 問中 27 問 (67.5%) である. 実際, 今回の実験で得られた最大正解数は 23 であり, 正解数 27 は, 現実的に到達可能な範囲にあると考えられる.

### 5.4 好成績の理由を考える

このような比較的単純な解法でも, 半数以上の設問が正しく解けるのは, どうしてだろうか. その理由は, おそらく, 「センター試験がよく練られた試験問題である」ということになるう.

<sup>8</sup> unidic では, 語彙素を原形として採用した.

<sup>9</sup> 事前選抜を上限の計算に含めるのは複雑なので, 除外した.

<sup>10</sup> 数字 84 は, 表 4-5 の行数の合計に, 4 は 1 行に対する設定数に, -4 は抽出法 P-0-0 は設定しないことに対応する.

<sup>11</sup> 表 7 の左側の区切り線は, この境界を示す.

表 7 各設問における正解順位分布 (「評論」)

設問	1 位	2 位	3 位	4 位	5 位	設問	1 位	2 位	3 位	4 位	5 位
2003S-E3	664	4	0	0	0	2003M-E4	202	103	8	34	321
2001S-E4	612	51	3	1	1	2011M-E3	195	159	105	80	129
2001M-E2	582	78	8	0	0	2005M-E4	181	156	133	130	68
2005M-E5	558	90	15	5	0	2011M-E5	177	290	149	42	10
2007S-E3	553	82	26	5	2	2003S-E4	174	143	171	148	32
2007M-E5	544	52	50	14	8	2001S-E2	172	268	189	36	3
2001M-E5	543	96	25	4	0	2007S-E5	168	386	79	32	3
2005S-E2	438	187	43	0	0	2007S-E4	123	353	173	19	0
2005S-E5	420	211	32	5	0	2005S-E3	102	25	74	212	255
2009M-E3	396	86	98	52	36	2005M-E3	100	150	248	107	63
2011M-E2	377	178	62	37	14	2003M-E2	77	279	47	265	0
2003S-E5	359	305	4	0	0	2005S-E4	76	74	79	156	283
2001S-E3	358	78	217	8	7	2007M-E4	45	66	267	207	83
2001M-E4	348	150	147	23	0	2001M-E3	44	102	219	198	105
2007S-E2	285	140	152	63	28	2009M-E2	25	117	236	286	4
2003M-E3	258	171	235	4	0	2011M-E4	20	151	290	193	14
2003S-E2	251	383	34	0	0	2009M-E5	14	197	120	254	83
2005M-E2	223	252	105	64	24	2003M-E5	14	28	86	124	416
2007M-E3	215	132	227	18	76	2001S-E5	0	3	37	31	597
2007M-E2	205	266	183	12	2	2009M-E4	0	1	14	151	502

センター試験の問題は、当然のことながら、「正解が一意に定まる（大多数の人が、正解に納得できる）」ことが必要である。

答の一意性を保証できる『数学』の問題とは異なり、『国語』の傍線部問題は、潜在的には多数の「正解文」が存在する。作問者の立場に立てば、そのうちの一つを選択肢に含め、それ以外を選択肢に含めないように問題を作らなければならない。そのため、正解選択肢とそれ以外の選択肢の間に、明示的な差異を持ち込まざるを得なくなる。そして、そのために持ち込まれた差異は、オーバーラップ率のような表層的な指標においても、識別できる差異として現れてしまうのであろう。もし、この推測が正しいとすれば、「良い問題であれば、機械にも解ける」ということであり、本論文で提案した解法は、センター試験ならではの性質を利用していることになる。

## 5.5 正解が得られない設問

すでに何度も述べたように、我々は、本論文で提案した解法で「評論」の傍線部問題の半数以上が解けてしまうことが驚きであり、解けない設問があることに何の不思議さも感じない。しかしながら、査読者より、採録条件として、「提案手法では正解が得られない設問に対する分析

（定性的な議論）が必要である」との指摘があったので、この点についての我々の見解を以下に述べる。

まず、（ある特定パラメータを使用した）この解法によって正解が得られない直接的な理由は、抽出した本文の領域  $\hat{T}$  と正解選択肢との照合スコア（オーバーラップ率）が低いことによる。この理由をさらに分解すると、次の3つの理由に行きつく。

**R1** そもそも本文中に根拠がない

**R2** 不適切な領域  $\hat{T}$  を抽出している

**R3** 照合スコアが意味的整合性を反映していない

しかしながら、特定の設問が解けない理由を、このどれか一つに特定することは難しい。

まず、理由 R1 であるが、確かに、これがほぼ明白なケースは存在する。たとえば、2003M-E4 は、傍線部の「具体例」を問う問題であるが、本文中にはその具体例は述べられていない。しかしながら、「根拠」という言葉はいささか曖昧であり、広くも狭くも解釈できるため、その解釈を固定しない限り、根拠の有無を明確に判断することは難しい。受験対策本がいう「根拠」は、「正解選択肢を選ぶ手がかり」という広い意味であり、「『適切に語句や表現を言い換えれば、選択肢の表現に変換できる』本文の一部」という狭い意味ではない。前者の意味では、ほとんどの設問に根拠は存在するが、後者の意味では、ほとんどの設問に根拠は存在しない。評論の傍線部問題で問われるのは、本文全体の理解に基づく傍線部の解釈であり、表現レベルの単純な言い換えではない。

次に、理由 R2 であるが、今回の解法では、選択肢と照合する領域として文または段落を単位とする連続領域のみを扱っている。しかし、実際の（広い意味での）根拠は、より小さな句や節といった単位の場合もあり、かつ、不連続に複数箇所存在する場合も多い。現在の実装の自由度における最適な領域が、必ずしも真の意味で適切な領域であるとは限らない。

最後に、理由 R3 であるが、現在の照合スコアが意味的整合性を適切に反映しない場合が存在するのは自明である。しかし、問題はそれほど単純ではない。照合スコアの具体的な値は、照合領域  $\hat{T}$  に依存する。最適な領域が定まれば、使用している照合スコア計算法の善し悪しを議論できるが、最適な領域が不定であれば、正解を導けない原因を、領域抽出の失敗 (R2) に帰すべきか、照合スコアの不適切さ (R3) に帰すべきかは、容易には定まらない。

以上のように、特定の設問が解けない理由を追求し、解けない設問を類形化することは、かなり難しい。さらに、チャンスレベルは 20% であるから、たまたま解ける設問も存在する<sup>12</sup>。そのような困難さを踏まえた上で、解けない設問を大胆に類形化するのであれば、次のようになる。

- 正解選択肢を選ぶ根拠が、傍線部のかなり後方に位置する設問。

<sup>12</sup> 前述の 2003M-E4 は正解する場合もある。



- 正解選択肢を選ぶ根拠が, 本文全体に点在している設問.
- 正解選択肢が, 本文全体の理解・解釈を前提として, 本文中には現れない表現で記述されている設問.
- 本文と整合しない部分を含む選択肢を除外していくこと (いわゆる消去法 (板野 2010)) によって, 正解選択肢が導ける設問.

## 5.6 「小説」に適用する

本論文で提案した解法を, そのまま「小説」の傍線部問題に適用すると, どのような結果が得られるであろうか. その疑問に答えるために, 「評論」と同様の実験を, 「小説」に対しても実施した. 対象とした「小説」の傍線部問題は計 38 問なので, ランダムに解答すると,  $7.6 \pm 4.83$  問 ( $p = 0.05$ ) の正解が得られることになる.

実験において, 統計的に有意な結果 (正解数 13 以上) が得られたのは, 2,004 通り中 13 通りの設定のみであった. これらを表 8 に示す. さらに, 「評論」と同様に, 各設問に対しても 668 通りの実験結果を集計した<sup>13</sup>. 正解数が  $133.6 \pm 20.26$  の上限を越えたのは, 38 問中 10 問であった. これらの結果より, 「小説」に対しては, 本論文で提案した解法の性能は, チャンスレベルと大差がないとみなすのが妥当であろう.

表 8 正解数が 13 以上の設定と正解の順位分布 (「小説」)

ID	bar	score	ps	ave.	1	2	3	4	5
S-1-0	bar	<i>W</i> -unidic	yes	1.2	15	6	4	7	6
S-2-0	bar	<i>L</i> -unidic	yes	2.2	15	4	2	11	6
P-0-0		$A^2$	no	6.0	14	5	6	7	6
P-0-0		$A^2$	yes	6.0	14	6	7	5	6
S-1-0	bar	<i>L</i> -unidic	yes	1.2	14	6	5	7	6
S-1-1	bar	<i>L</i> -unidic	yes	2.2	14	4	7	7	6
S-1-0		$A^2$	no	2.3	14	4	7	7	6
S-1-0	bar	$A^2$	yes	1.2	13	7	3	9	6
S-1-0	bar	<i>W</i> -unidic	no	1.2	13	6	5	6	8
S-1-1	bar	<i>W</i> -unidic	yes	2.2	13	8	4	7	6
S-3-1	bar	$A^2$	yes	4.2	13	5	7	7	6
S-3-2	bar	$A^2$	yes	5.2	13	5	6	8	6

<sup>13</sup> 表 8 の結果に基づき, 形態素解析辞書には ipadic ではなく unidic を採用した.

## 6 結論

本研究で得られた結果をまとめると、次のようになる。

- (1) センター試験『国語』現代文の傍線部問題に対する解法を提案・定式化した。
- (2) 本論文で示した解法は、「評論」の傍線部問題に対しては有効に機能し、半数以上の設問に対して正解を出力することができる。今回の実験から推測される性能の上限は95%、現実的に到達可能な性能は65-70%である。
- (3) 本論文で示した解法は、「小説」の傍線部問題に対しては機能しない。その性能はチャンネルと同等である。

本論文で示した解法で「『評論』が解ける」という事実を言い換えるならば、それは、「『評論』では、本文に書かれていることが問われる」ということである。これに対して、「『小説』が解けない」という事実は、その裏返し、すなわち、「『小説』では、本文に書かれていないこと（心情や行間）が問われる」ということを示している。このような差異の存在を、船口（船口 1997）も指摘しているが、表層的なオーバーラップ率を用いる比較的単純な方法においても、その差異が明確な形で現れることが判明した。

## 謝 辞

本研究では、国立情報学研究所のプロジェクト「ロボットは東大に入れるか」から、データの提供を受けて実施した。本研究の一部は、JSPS 科研費 24300052 の助成を受けて実施した。本研究では、mecab/ipadic、mecab/unidic を使用した。

## 参考文献

- 新井紀子, 松崎拓也 (2012). ロボットは東大に入れるか?—国立情報学研究所「人工頭脳」プロジェクト—. 人工知能学会誌, **27** (5), pp. 463-469.
- 船口明 (1997). きめる! センター国語現代文. 学研教育出版.
- 服部昇平, 佐藤理史 (2013). 多段階戦略に基づくテキストの意味関係認識: RITE2 タスクへの適用 情報処理学会研究報告, 2013-NLP-211 No.4/2013-SLP-96 No.4, 情報処理学会.
- Hattori, S. and Sato, S. (2013). "Team SKL's Strategy and Experience in RITE2" In *Proceedings of the 10th NTCIR Conference*, pp. 435-442.
- 板野博行 (2010). ゴロゴ板野のセンター現代文解法パターン集. 星雲社.
- 教学社編集部 (2013). センター試験過去問研究 国語 (2014 年版センター赤本シリーズ). 教学社.

## 略歴

**佐藤 理史**：1988 年京都大学大学院工学研究科博士後期課程電気工学第二専攻研究指導認定退学。京都大学工学部助手，北陸先端科学技術大学院大学助教授，京都大学大学院情報学研究科助教授を経て，2005 年より名古屋大学大学院工学研究科教授。工学博士。現在，本学会理事。

**加納 隼人**：2010 年名古屋大学工学部電気電子・情報工学科入学。2014 年同学科卒業。現在，名古屋大学大学院工学研究科電子情報システム専攻在学中。

**西村 翔平**：2010 年名古屋大学工学部電気電子・情報工学科入学。2014 年同学科卒業。

**駒谷 和範**：1998 年京都大学工学部情報工学科卒業。2000 年同大学院情報学研究科知能情報学専攻修士課程修了。2002 年同大学院博士後期課程修了。博士（情報学）。京都大学大学院情報学研究科助手・助教，名古屋大学大学院工学研究科准教授を経て，2014 年より大阪大学産業科学研究所教授。現在，SIGDIAL Scientific Advisory Committee member。情報処理学会，電子情報通信学会，人工知能学会，言語処理学会各会員。

(2013 年 9 月 10 日 受付)

(2013 年 11 月 5 日 再受付)

(2014 年 1 月 15 日 再々受付)

(2014 年 1 月 30 日 採録)